

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: 10319948 A

(43) Date of publication of application: 04 . 12 . 98

(51) Int. Cl

G10H 1/00
G01R 23/16
G10L 3/00

(21) Application number: 09125864

(71) Applicant: NIPPON TELEGR & TELEPH CORP <NTT>

(22) Date of filing: 15 . 05 . 97

(72) Inventor: KAYANO KUNIO MURASE HIROSHI

(54) SOUND SOURCE KIND DISCRIMINATING METHOD OF MUSICAL INSTRUMENT INCLUDED IN MUSICAL PLAYING

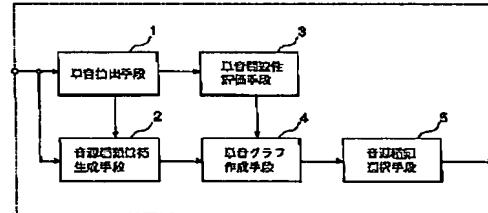
sound source kind candidate is changed by a probabilistic transmission operation (5).

COPYRIGHT: (C)1998,JPO

(57) Abstract:

PROBLEM TO BE SOLVED: To precisely conduct a discrimination even when many monotones are included or the characteristics of the monotones are greatly varied locally.

SOLUTION: A monotone is extracted from inputted acoustic signals (1). Then, the candidates for the sound source kind (names of musical instruments) of each monotone and the associated degrees of confidence level are obtained (2). Then, the relationship between two different monotones, which are extracted, is evaluated employing the transition probability of a pitch, the degree of similarity of a tone color, the proximity in timing and the characteristic of relative pitch in the playing (3). Then, the monotone graph of a Bayesian network is generated to express the monotones having a strong relationship as a link while making a monotone as a node based on the sound source candidates of the obtained monotones and the degrees of confidence levels (4). Then, by utilizing the fact that the monotones having a strong relationship on the monotone graph are played by a same musical instrument kind with a high probability, a most probable sound source kind candidate is selectively outputted after the credibility degree of



BEST AVAILABLE COPY

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平10-319948

(43)公開日 平成10年(1998)12月4日

(51)Int.Cl.[®]

G 10 H 1/00
G 01 R 23/16
G 10 L 3/00

識別記号

531

F I

G 10 H 1/00
G 01 R 23/16
G 10 L 3/00

B
B
531 N

審査請求 未請求 請求項の数 3 OL (全 11 頁)

(21)出願番号

特願平9-125864

(22)出願日

平成9年(1997)5月15日

(71)出願人

日本電信電話株式会社

東京都新宿区西新宿三丁目19番2号

(72)発明者

柏原 邦夫
東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(72)発明者

村瀬 洋
東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(74)代理人 弁理士 草野 卓

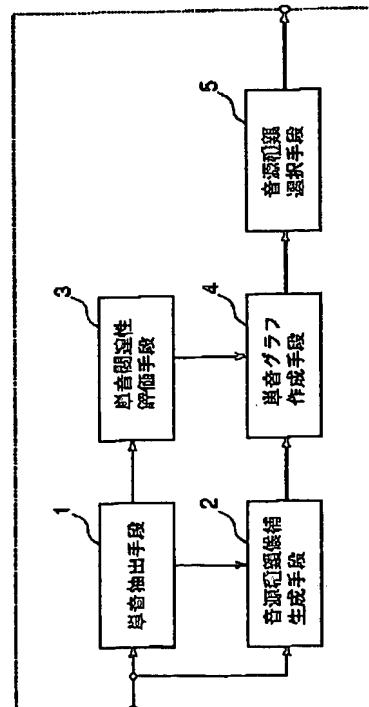
BEST AVAILABLE COPY

(54)【発明の名称】 音楽演奏に含まれる楽器音の音源種類判別方法

(57)【要約】

【課題】 同時に多くの単音が含まれていても、また単音の性質が局所的に大きく変動しても高い精度で判別できる。

【解決手段】 入力音響信号から単音を抽出し(1)、抽出された各単音の音源種類(楽器名)候補とその確信度を求め(2)、抽出された単音中の異なる二つの単音どうしの関連性を、音高の遷移確率、音色の類似性、時間の近接性、演奏における相対的音高の性質の観点から評価し(3)、求めた単音の音源種類候補とその確信度、および二つの単音の関連性に基づいて、単音をノードとして関連性の高い単音どうしをリンクとして表現したペイジアンネットワークの単音グラフを作成し(4)、その単音グラフ上で関連性の強い単音は同一楽器種類で演奏される可能性が大きいという性質を利用して、確率伝搬操作によって音源種類候補確信度を変更した後、最も確信度の高い音源種類候補を選択出力する。



【特許請求の範囲】

【請求項1】 音楽演奏の音響信号に含まれている単音を抽出する単音抽出過程と、上記抽出された単音の音源種類（単音を演奏している楽器名）の候補を、その各候補に対する確信度とともにに出力する音源種類候補生成過程と、一連の音楽演奏における、旋律など、単音どうしの関連性を上記抽出された単音について評価する単音関連性評価過程と、上記音源種類候補生成過程および上記単音関連性評価過程の各出力に基づいて、単音をノードとし、関連性の高い単音どうしをリンクとして表現したグラフを作成する単音グラフ作成過程と、この単音グラフ作成過程によって作成された単音グラフ上で、関連性の強い単音は同一の楽器種類で演奏される可能性が大きいという音楽の一般的な性質を利用して、上記音源種類候補生成過程によって出力された音源種類候補に対する確信度を変更した上で、最も確信度の高い音源種類候補を選択して出力する音源種類選択過程とを有することを特徴とする、音楽演奏に含まれる楽器音の音源種類判別方法。

【請求項2】 上記単音関連性評価過程は、異なる二つの単音どうしの関連性を、音高の遷移確率、音色の類似性、時間の近接性、および演奏において同じ役割を果たしているかどうかの、少くとも一つについてのそれぞれの評価値を用いて評価することを特徴とする請求項1記載の音楽演奏に含まれる楽器音の音源種類判別方法。

【請求項3】 上記単音グラフ作成過程は、単音をノードとし、関連性の高い単音どうしをリンクとして、確率値の伝搬によって確信度を変更することのできるペイジアンネットワークを作成することを特徴とする請求項1または2記載の音楽演奏に含まれる楽器音の音源種類判別方法。

【発明の詳細な説明】**【0001】**

【発明の属する技術分野】 この発明は、複数種類の楽器による音楽演奏に対する自動採譜装置、採譜支援装置、および音楽演奏の記号化装置などの構成要素として用いられる、音楽演奏に含まれる楽器音の音源種類を判別する方法に関する。

【0002】

【従来の技術】 従来、音楽演奏に含まれる楽器音の音源種類判別方法に関しては、各楽器音のパワースペクトルの特徴に着目して、各単音ごとに判別分析を行って楽器音を判別する方法が知られている。しかし、この方法では、一般の楽器演奏にしばしば見られるように、同時に数多くの単音が発音している場合には、各単音に由来するパワースペクトルが重なり合うことによって安定した特徴の抽出が困難となるため、精度の高い判別分析が行い難いという欠点があった。

【0003】 また、予め対象とする楽器音の音響信号波形またはスペクトログラムを蓄積しておき、入力音響信号を所定の時間窓（フレーム）ごとに区分した上で、各区分に対して、蓄積したその音響信号波形またはスペクトログラムと照合を行い、最もよく当てはまる音響信号波形またはスペクトログラムを選択することによって楽器音を判別する方法が知られている。しかし、この方法では、ある時間区分ごとに処理が行われ、旋律、ベースラインなどといった、時間をまたがった単音のつながりの情報が考慮されていないため、例えばフルートによって演奏されているメロディーの流れの中で、フルートと認識されるべき単音が突然トランペットの音と誤認識される例があるなど、特に、表情豊かな演奏に見られるような単音の特徴が大きく変動する演奏や、比較的同時発音数の多い演奏などにおいて、音楽的に不自然な楽器音判定誤りが生じることが多いという欠点があった。

【0004】 従って、上記の各方法は、同時に数多くの単音が発音している場合や、一連の演奏において単音の性質が局所的に大きく変動するような場合にあっては、十分に高精度な、音楽演奏に含まれる楽器音の音源種類判別処理が期待し難い。

【0005】

【発明が解決しようとする課題】 この発明は、同時に数多くの単音が発音している場合や、一連の演奏において単音の性質が局所的に大きく変動するような場合であっても適用することのできる、音楽演奏に含まれる楽器音の音源種類判別の処理方法を提供し、従来の方法と比較して音楽演奏に含まれる楽器音の音源種類判別処理を高い精度で行うことを目的としている。

【0006】

【課題を解決するための手段】 この発明によれば、音楽演奏の音響信号に含まれている単音を単音抽出過程で抽出し、その抽出した単音の音源種類、すなわちその単音を演奏している楽器名が何であるかの候補を、その各候補に対する確信度とともに音源種類候補生成過程で求め、一連の音楽演奏において、旋律など、単音どうしの関連性を、抽出した単音について単音関連性評価過程で評価し、音源種類候補生成過程により求めた音源種類候補とその確信度および単音関連性評価過程で求めた単音の関連性に基づいて、単音をノードとし、関連性の高い単音どうしをリンクとして表現したグラフを単音グラフ作成過程で作成し、この単音グラフ作成過程によって作成された単音グラフ上で、関連性の強い単音は同一の楽器種類で演奏される可能性が大きいという音楽の一般的な性質を利用して、音源種類候補生成過程によって求めた音源種類候補に対する確信度を変更した上で、最も確信度の高い音源種類候補を選択して出力することを音源種類選択過程で行う。

【0007】

【発明の実施の形態】 次に、この発明の実施形態について

て図面を用いて説明する。図1は、この発明方法を適用した音楽演奏に含まれる楽器音の音源種類判別装置の一実施形態を示すブロック図である。この音楽演奏に含まれる楽器音の音源種類判別装置は、単音抽出手段1と、音源種類候補生成手段2と、単音関連性評価手段3と、単音グラフ作成手段4と、音源種類選択手段5で構成され、音楽音響信号を入力とし、演奏に含まれる単音の発音時刻、音高、および楽器種類を表した記号の列を出力する。

【0008】単音抽出手段1は、音楽演奏の音響信号を入力とし、周波数解析、周波数成分の抽出、および単音の立上り検出を行うことによって単音を抽出する。音源種類候補生成手段2は、単音抽出手段1の出力する単音の音響信号波形を、予め蓄積してある対象とする楽器音の単音の音響信号波形と照合することによって、音源種類の候補を生成する。

【0009】単音関連性評価手段3は、単音抽出手段1の出力する単音を入力とし、異なる二つの単音どうしの関連性を、音高の遷移確率、音色の類似性、時間の近接性、および演奏における相対的音高の性質の観点から評価する。単音グラフ作成手段4は、音源種類候補生成手段2および単音関連性評価手段3の各出力に基づいて、単音をノードとし、関連性の高い単音どうしをリンクとして表現したグラフ（ペイジアンネットワーク）を作成する。

【0010】音源種類選択手段5は、単音グラフ作成手段4で出力される単音グラフ上で、関連性の強い単音は同一の楽器種類で演奏される可能性が大きいという音楽の一般的な性質を利用して、ペイジアンネットワーク上における確率伝搬操作によって音源種類候補生成手段2で出力された音源名候補に対する確信度を変更した上で、最も確信度の高い音源種類候補を選択して出力する。

【0011】次に、上述した各手段1、2、3、4、5における各処理の流れを、図2から図6にそれぞれ示す流れ図を参照して具体的に説明する。単音抽出手段1では、まず装置への入力となる音楽演奏の音響信号波形を読み込む（ステップ101）。音響信号波形の例を図7に示す。図7は、ある曲をバイオリン、フルート、およびピアノを用いてアンサンブル演奏した音響信号の例である。次に、読み込んだ波形に対し周波数解析を行って、スペクトrogramを得る（ステップ102）。スペクトrogramは、音響信号に含まれるパワーを、横軸時間、縦軸周波数の平面上に表現したものである。スペクトrogramの例を図8に示す。図8は、中心周波数の異なる多数の帯域通過フィルタを用いて得たものである。図8において、各時点上の縦線の途中が右へ山状にわん曲している部分は、その縦軸における周波数成分のパワーの大きさを表す。続いて、ステップ102で得たスペクトrogramに対して、周波数成分を抽出する（ステップ103）。周波数成分とは、スペクトrogram上にお

ける一連のローカルピークのことである。周波数成分抽出は、スペクトrogramをまず周波数方向に走査してパワーのローカルピークを検出し、このローカルピークを時間方向に接続することによって行う。抽出された周波数成分の例を図9に示す。図9では、周波数成分が線分で表されており、その立上りに○印がつけられている。このような周波数成分がいくつか集まって単音が形成されるが、音楽演奏においては、打楽器音などの例外を除けば、同じ単音を形成する周波数成分は、立上り時刻が近接しており、また周波数が整数倍の関係にあることが知られている。そこで、各周波数成分の立上り時刻の近接性に基づいて、同時に発音している単音を形成する周波数成分をひとまとめにする（ステップ104）。それぞれのまとめには、複数の単音が含まれている可能性があるので、それぞれのまとめにおいて、周波数成分の周波数を低い周波数から順に調べ、整数倍の関係にない成分を抽出する。これによって、単音の基本周波数を得ることができる（ステップ105）。このようにして、単音の基本周波数、立上り時刻、および立下り時刻が得られるので、これらを単音の情報として出力する（ステップ106）。この段階では、まだそれぞれの単音がいかなる種類の音源であるかを判別する処理は行われていない。なお図9における各○印の所に付けた数字の左の値は、発生時刻対応に付けた番号であり、右の値は同一時刻内で付けた番号である。

【0012】音源種類候補生成手段2（図3）では、まず単音抽出手段1で出力された単音抽出結果を読み込む（ステップ201）。次に、それぞれの単音に相当する時間区間の入力音響信号について、これに含まれている可能性のある単音の記憶波形（以下、候補波形という）を、あらかじめ蓄積した音響信号波形の中から選択する（ステップ202）。この選択は例えば次のように行う。単音に相当する時間区間の入力音響信号について先に抽出した周波数成分を用いて基本周波数およびパワー包絡等の音の特徴量を抽出する。この特徴量は、その入力音響信号に含まれている可能性のある音の記憶波形を選択するために用いられる。音の記憶波形は、図に示していない波形記憶手段にあらかじめ蓄積されているので、これを順に検査する。即ち各記憶波形ごとにその基本周波数と、入力音響信号の抽出された周波数成分の周波数とを比較し、ある範囲内に収まっているかどうかを調べ、もしある範囲に収まっているければ、その記憶波形は当該入力音響信号に含まれている可能性は低いので、その記憶波形は候補から外す。前記ある範囲は例えば次のようにして決める。即ち記憶波形の基本周波数をその大きさ順に並べた場合、ある基本周波数についてみると、そのすぐ下の基本周波数との間の半分だけ低い周波数から、すぐ上の基本周波数との間の半分だけ高い周波数までの範囲に入るものを候補とする。例えば半音ごとの記憶波形を設ける場合は、半音は6%ずつ周波数が

高くなっているから、基本周波数±3%の範囲にあるものを候補とする。前記ある範囲に収まつていれば、さらに特徴量に矛盾（例えば発音不可能な音域であるなど）があるかどうかを調べ、もし矛盾があれば、その記憶波形は候補から外す。続いて、それぞれの単音に相当する時間区間の入力音響信号と、ステップ202で選択した候補波形との照合を行う（ステップ203）。この照合は、単音の多様性や変動を吸収するために候補波形を入力音響信号波形に適応させる処理を含めて特願平9-3

$$y_n(k) = \sum_{m=0}^{M-1} h_n(m) r_n(k-m)$$

と書ける。ここでkは標本化された時刻、nは候補波形を数える添字、 $y_n(k)$ はこのフィルタ演算を適用した結果の波形、 h はFIRフィルタのインパルス応答、 r

$$J = E [\{ z(k) - \sum_{n=0}^{N-1} y_n(k) \}^2]$$

と書ける。ここで $z(k)$ は区分入力音響信号波形、Nは候補波形の数、Eは時間平均を表す。これを最小化するための必要条件は、全てのnとmに関して、偏微分 ∂J

$$\begin{aligned} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} E [r_i(k-j) r_n(k-m)] h_n(m) \\ = E [r_i(k-m) z(k)] \end{aligned}$$

を導くことができる。方程式(3)を解くことによって、入力音響信号を、最も少ない誤差で解釈するための、各候補波形に対するフィルタ係数を求めることができる。各候補波形に対して式(1)のフィルタ演算を行った結果得られる波形の平均パワーが大きいと、その候補波形が入力音響信号中に大きい割合で含まれていると判断できる。そこで、ステップ202で選択された候補波形の種類を音源種類候補とし、上記フィルタ演算の結果の平均パワーを、その音源種類が入力音響信号中に含まれている確信度として出力する（ステップ204）。

【0014】図10は、音源種類候補生成手段2によって生成された、音源種類候補の情報をともなった単音を示したものである。図は、横方向が時間を、また縦方向が単音の音高を表す。それぞれの四角形が単音を表す。四角形の上部の数字は、単音の通し番号と、音高を表す番号である（例えば11-58は、音高番号が58番の、通し番号11番の単音である）。それぞれの四角形に表示されているバーグラフは、上位3位までの音源種類に対する確信度を示している。音源種類がバイオリンはVで、フルートはFで、ピアノはPでそれぞれ示している。

【0015】単音関連性評価手段3、単音グラフ作成手段4、および音源種類選択手段5が、この発明の中心をなす部分である。単音関連性評価手段3（図4）では、はじめに、単音抽出手段1で得た単音抽出結果を読み込み（ステップ301）、それぞれの単音を一つずつ選択しながら処理を行う。まず、選択した単音（これを単音

$$D = \{(0.8-0.7)^2 + (0.3-0.8)^2 + (0.1-0.3)^2\}^{1/2} \quad (4)$$

この距離の値Dを、予め一連の旋律中に現れる単音についてどの程度の距離Dをもつた単音が現れるかの統計（ヒストグラム）をとっておけば、(4)式で求められ

1813「音響信号分離方法」の明細書中で述べた方法を用いて行うことができる。

【0013】即ち、候補波形に対して適応のためのフィルタ演算を適用した結果の波形を足し合わせた波形と、単音に相当する時間区間の入力音響信号との平均自乗誤差が最小となるようなフィルタ係数を求めるために、連立方程式を作成する。フィルタとしてFIR型を用いることにすれば、候補波形にフィルタ演算を適用した結果の波形は

$$y_n(k-m) \quad (1)$$

は候補波形、Mはフィルタの次数である。すると、候補波形にフィルタ演算を適用した結果の波形を足し合わせた波形と、この区分入力音響信号との平均自乗誤差は

$$\sum_{n=0}^{N-1} \sum_{m=0}^{M-1} E [\{ z(k) - y_n(k-m) \}^2] \quad (2)$$

$\partial h_n(m)$ が0となることである。この条件を用いると、 $N \times M$ 個の連立一次方程式

$$\sum_{n=0}^{N-1} \sum_{m=0}^{M-1} E [r_i(k-j) r_n(k-m)] h_n(m) = E [r_i(k-m) z(k)] \quad (3)$$

Aとよぶ）に対して、その読み込んだ単音中に関連性のある単音があるかどうかを検索する（ステップ302）。もしなければ、次の単音を選択するが、選択できる単音がなければ処理を終了する。もし単音Aと関連性のある単音が検索されれば、その単音（これを単音Bとよぶ）と単音Aとの関連性の高さを評価する。このことは逆に単音Aから単音Bへの遷移の現れにくさによっても評価することができる。つまり遷移の現れにくさが小さい程、関連性が高いことになり、この評価は次のようにして行うことができる。なお、以下で、単音Bは単音Aよりも発音開始時刻が前であるとする。

【0016】まず、単音Bと単音Aの音色の非類似性 P_1 を評価する（ステップ303）。これは、音源種類候補生成手段2で得た確信度をベクトルと見て、そのノルムの値を用いて行う。確信度が0～1までの範囲の値である方が良いので、例えば、単音Bの確信度が

フルート 0.8

バイオリン 0.3

ピアノ 0.1

であり、単音Aの確信度が

フルート 0.7

バイオリン 0.8

ピアノ 0.3

であったならば、音色間の距離Dは、次式で求められる。

【0017】

$$D = \{(0.8-0.7)^2 + (0.3-0.8)^2 + (0.1-0.3)^2\}^{1/2} \quad (4)$$

た距離の値を、確率値に変換することができる。この確率値を P_1 として単音Bと単音Aの類似性の評価値とする。

【0018】次に、単音Bと単音Aの音高の関係を評価する（ステップ304）。これは、予め対象とする楽曲のジャンル（ジャズ、ポップスなど）において、音高の遷移の統計をとっておき、この統計データに基づいて行う。例えば、単音Bの音高がC4（「中央ド」）であり、単音Aの音高がE4（そのすぐ上の「ミ」）であつ

$$W(\delta t) = \exp((t_B - t_A)/\tau)$$

τ は時定数であり、Wは単音A、B間の遷移、つまり「音の流れやすさ」の評価尺度における時間窓に相当する。この値が小さいほど、単音Bと単音Aの時間的な関係が密接であることを表す。

【0020】次に、単音Bと単音Aの役割を評価する（ステップ306）。ここで、単音の役割とは、主旋律（アンサンブル演奏の場合、その時刻において最高音となることが多い）、ベースライン（同、最低音となることが多い）などのような、演奏における音楽的な役割のことである。ある役割をもつ音は、同じ役割をもつ音との連続性が大きいと考えられる。この役割の評価も遷移の現われやすさとして次のように確率で表すことができる。

【0021】

$P_3(n_k, n_{k-1}) = P(S_k | S_{k-1})$
 S_k はk番目の単音の役割を表し、これを単音の役割可担性と呼ぶ、例えば $P(S_k | S_{k-1})$ は n_{k-1} が最高（最低）音であったとき n_k も最高（最低）音である確率を表す。この役割可担性は次式のようにN-gramの形を用いて表すこともできる。

【0022】

$P_3 = P(S_k | S_{k-1}, S_{k-2}, \dots, S_{k-N+1})$
 $S_{k-1}, S_{k-2}, \dots, S_{k-N+1}$ はそれぞれ最高音又は最低音であり、この系列の次の S_k が最高（最低）音である確率が P_3 である。 P_3 は統計的分析によって得ることができるが、次の実験式として求めることもできる。

【0023】 $P_3 = aq + b$

a, bは定数で、qは既存の単音の内の音の流れ（music stream）上の最大10個前までの単音に占める最高（最低）音であった単音の割合である。次に、ステップ303～ステップ306で得られた評価値を統合して出力する（ステップ307）。評価値の統合は、それぞれの評価値の対数の重みつきの和

$Z(A, B) = W \sum_i \{-w_i \log P_i(A, B)\}$
 を計算することによって行う。この評価値が小さいほど、単音Bと単音Aの関連性が高いことを表す。続いて、ステップ302に戻って、次の単音の組合せを選択して、関連する可能性のある全ての組合せを評価するまで処理を繰り返す。

【0024】単音グラフ作成手段4（図5）では、単音をノードとし、いわゆる最急降下法によって、局所的に最も関連性の高い単音どうし、つまり前記Zが最小となる方向に順次単音をリンクで結びつけて、時間を追って

たならば、統計データを参照してドからミへの遷移の現れる確率 P_2 を得て、音高の関係の評価値とする。

【0019】次に、単音Bと単音Aの時間関係を評価する（ステップ305）。すなわち、単音Bの終了時刻を t_B とし、単音Aの開始時刻を t_A とすると、次式を時間関係の評価値とする。

$$(5)$$

順次グラフを生成する処理を行う。まず、音源種類候補生成手段2で生成された、音源種類候補の情報をともなった単音のうちで、未処理の単音を検索し（ステップ401）、未処理の単音がなければ処理を終了する。未処理の単音があれば、その単音（これを単音Aとする）のノードを作成する（ステップ402）。次に、単音関連性評価手段3の処理結果を用いて、既に作成した単音ノード（単音Aよりも早い開始時刻をもつ単音に対応する）の中から、単音Aと最大の関連（Zが最小）をもつ単音（これを単音Bとする）を選択する（ステップ403）。続いて、単音Bから見て単音Aが最大の関連（Zが最小）をもつかどうかを調べ（ステップ404）、もしもたないならば、ステップ401に戻り、もしもつならば、単音Bに既に下位の（時間的に後の）リンクが存在しているかどうかを調べる（ステップ405）。もしなければ、直ちに、またもし既にリンクがあれば、これを切断した上で（ステップ406）、単音Bのノードから単音Aのノードにリンクを作成して（ステップ407）、ステップ401に戻る。このようにして作成されたグラフは、強い関連をもつ単音のつながりをグラフの構造として表現したものである。

【0025】図12に示すように、新ノード n_k から既存ノードnのリンク候補 $l_1 \sim l_4$ のうちで、最大の関連があるもの l_1 を選択し、そのリンク l_1 で接続されるノード n_{k-3} からのリンク候補 $g_1 \sim g_3$ のうちで最も関連があるもの g_1 を選択し、 $l_1 = g_1$ であれば、それをリンクとし、ノード n_{k-3} からリンク g_1 以外の方向にリンクが存在していれば、そのリンクは切断する。

【0026】音源種類選択手段5では、単音グラフ作成手段4で作成されたグラフを用いて、単音の音源種類確信度を更新した上で、確信度最大の音源種類を選択する。音源種類確信度の更新の基礎となる原理はペイズの定理に基づくものであり、原理としては公知のもの（例えば、柏野邦夫ほか：「音楽情景分析の処理モデルOPTIMAにおける単音の認識」、電子情報通信学会論文誌、J79-D-II、pp. 1751-1761、1996年11月）である。この原理に基づいて確信度の更新を行うグラフは、ペイジアンネットワークと呼ばれる。

【0027】ペイジアンネットワークによれば、各ノードにおける確信度ベクトル（ここでは、音源種類候補に対する確信度）と、リンクで結ばれたノード間の条件付

確率とを与えることによって、ノードの確信度ベクトルを、リンクを考慮した上での値に更新することができる。これは、各ノードにおける確信度ベクトルを、そのノードよりも上位（ここでは、時間的に前）のノードによってもたらされるそのノードへの影響 π と、そのノードよりも下位のノードによってもたらされるそのノードへの影響入との積で与えられることと、入と π とが、グラフ上をそれぞれ上位方向・下位方向にたどりながら順次定めることができるという性質によるものである。

【0028】単音グラフ作成手段4（図6）においては、まず、下位方向から上位方向にたどりながら、各ノ

$$\lambda_A \text{ (フルート) } = 0.8$$

$$\lambda_A \text{ (ピアノ) } = 0.2$$

が、またノードBには、

$$\lambda_B \text{ (フルート) } = 0.7$$

$$\lambda_A \text{ (ピアノ) } = 0.2$$

$$\lambda_B \text{ (バイオリン) } = 0.1$$

と与えられているとする。さらに、リンクで結ばれている単音が同じ音源種類で演奏されている確率を0.9、違う音源種類で演奏されている確率を0.1とする（これらの値は、単音関連性評価過程で計算される単音どう

$$\lambda_A' (n) = \alpha \sum_m p \lambda_A(n) \lambda_B(m)$$

$$n=m \text{ で } p=0.9, n \neq m \text{ で } p=0.1$$

と与えられるので（ α は正規化定数）、

$$\lambda_A' \text{ (フルート) } = 0.910$$

$$\lambda_A' \text{ (ピアノ) } = 0.090$$

となる（ステップ501）。次に、ノードAの π_A を用いてノードBの π_B を更新する。更新式は、ペイジアンネットワークにおける情報伝搬法より、ノードBのほか

$$\pi_B' (n) = \beta \sum_m p \pi_B(m)$$

$$n=m \text{ で } p=0.9, n \neq m \text{ で } p=0.1$$

と与えられる（ β は正規化定数）。この例では、 π_A は等確率と仮定すると、単に π_B も等確率となる（ステッ

$$\pi_A \text{ (フルート) } = 0.910$$

$$\pi_A \text{ (ピアノ) } = 0.090$$

と計算でき（ステップ503）、フルートの可能性が高いノードBと関連が強いという情報によって、ノードAがフルートの単音である可能性が、以前に増して強まったことが分かる。なお図10の単音ノードに対し、単音からの作成処理及び音源種類仮説選択処理中のステップ503までの処理を行った単音グラフを図11に示す。

【0030】前記評価値の統合した値Zにおける各評価値の重み w_i としては例えば音色類似性に対しては $w_1 = 1.0$ 、音高遷移性に対しては $w_2 = 0.35$ 、役割可担性については $w_3 = 1.0$ 、 $a = 0.8$ 、 $b = 0.1$ 時間関係については $\tau = 2$ 秒とそれぞれ選定される。なおこれら評価値は少くとも1つ用いればよく、この場合、前記重み w_i が小さいもの程、省略可能である。関連性評価をZで定義して、Zが最小となるように単音を順次つないだが、関連性が高いものを順次つなぐよう

一ノードにおける入値を計算する（ステップ501）。次に、グラフを上位方向から下位方向にたどりながら、各ノードにおける π 値を計算する（ステップ502）。この後、各ノードにおいて、入値と π 値との積を計算することにより、各ノードでの確信度が求められる（ステップ503）。最後に、各ノードで確信度最大の音源種類候補を出力する（ステップ504）。

【0029】ステップ501から503を簡単な例で説明する。ノードAとノードBが、この順にリンクで結ばれており、ノードAには、音源種類候補と確信度が

(6)

(7)

(8)

(9)

(10)

しの関連性の強さによって変えることも考えられる）。このとき、ステップ501では、下位ノードであるノードBの λ_A を用いてノードAの λ_A を更新する。更新式は、ペイジアンネットワークにおける情報伝搬法より、

(11)

(12)

(13)

(14)

にはノードAから直接リンクのつながっているノードがない場合には、

(15)

(16)

（ステップ502）。結局、ノードAの音源種類候補における確信度Pを計算すると、 π と入の積より、

(17)

(18)

関連性評価をしてもよい。

【0031】

【発明の効果】次に、この発明を適用した装置の認識精度を評価する実験について述べる。実験では、テスト曲として3パートのアンサンブルにアレンジした図14に示す曲を用いた。その曲を高音のパートから順にバイオリン、フルート、およびピアノで演奏し、スタジオでモノラル収録して実験用の音響信号試料とした。

【0032】この実験では音源同定処理の精度を測るために、音高と時刻については人手で正解を与えて実験した。実験に用いたパラメータ値 $w_1 \sim w_3$ 、 τ 、 a 、 b は先に例として上げたものを用いた。音源同定精度Rの定義は

$$R = 100 \left((\text{right} - \text{wrong}) / \text{total} \cdot 1/2 + 1/2 \right) (%)$$

とした。rightは出力に含まれる音符のうち音高と音色の両方が正しく認識された音符の数、wrongは出力に含まれる音符のうち、音高と音色のどちらか又は両方が正しくない音符の数、totalは入力（正解）に含まれる総音符数である。テンプレートフィルタリングのタップ数は40とした。テスト曲の演奏は、テンプレートの蓄積に用いたものとは別の各楽器個体を用いた。

【0033】図10は、既に述べたように、音源種類候補生成手段2によって出力された音源種類候補である。また図11は、音源種類選択手段のステップ503までの処理を行った結果である。図10と図11を比較すると、バイオリンのパート、フルートのパート、ピアノのパートが正しくリンクによって結ばれ、確信度の更新によって、各ノードにおいてはそれぞれのパートに対応する音源種類の確信度が最大となっていることが分かる。

【0034】図13に実験結果の同定精度を、図15に記載結果を楽譜形式でそれぞれ示し、図15Aは従来法の結果、図15Bはこの発明方法の結果であり、また音価の同定処理は行わず、全て四分音符として実時間上に示している。図15、A、Bを比較すると、この発明方法によればいくつかの音符の同定誤りが正しく修正されていることがわかる。

【0035】また図13によれば単音関連性評価手段3で用いた3つの要因は、音源同定精度の改善に対していずれも効果的であり、3つ全てを用いた場合には、これら関連情報を用いない場合に比べ、誤りをおよそ2/3に減少させることができた。また音高遷移性と音色類似性については単独で用いただけでは、音源同定精度に対してはかえって低下させる結果となっている。この図14からは役割可担性のみを用いてもよく、また、音高遷移性と音色類似性の組合せ、音高遷移性と役割可担性の組合せ、音色類似性と役割可担性の組合せの順に音源同定精度がよくなっている。全体を用いると、関連性情報

（単音間の音の流れ）を考慮しない場合は66.3%であった音源同定精度が、関連情報を考慮したこの発明によれば78.9%まで向上した。

【0036】実験の結果をまとめると、音源種類判定の正解率は関連性情報を適用しない場合は66.3[%]であるが、関連性情報を適用した場合は78.9[%]に向上了。関連性情報の適用、すなわちこの発明の中

心部分である単音関連性評価過程、単音グラフ作成過程、および音源種類選択過程の処理により、音源種類判定の正解率が向上していることが分かる。

【0037】以上、説明したように、この発明によれば、音楽演奏における単音の音楽的なつながりを考慮して楽器種類の判別を行うことによって、同時に数多くの単音が発音している場合や、一連の演奏において単音の性質が局所的に大きく変動するような場合であって、従来の方法に比較して高い精度で音楽演奏に含まれる楽器音の音源種類判別処理を行うことができる。

【図面の簡単な説明】

【図1】この発明による音楽演奏に含まれる楽器音の音源種類判別方法を適用した機能構成を示すブロック図。

【図2】図1中の単音抽出手段1の処理手順を示す流れ図。

【図3】図1中の音源種類候補生成手段2の処理手順を示す流れ図。

【図4】図1中の単音関連性評価手段3の処理手順を示す流れ図。

【図5】図1中の単音グラフ作成手段4の処理手順を示す流れ図。

【図6】図1中の音源種類選択手段5の処理手順を示す流れ図。

【図7】入力音響信号波形の一例を示す図。

【図8】図7の波形に対して周波数解析を行った結果得られたスペクトログラムを示す図。

【図9】図8のスペクトログラムに対して周波数成分抽出処理を行った結果を示す図。

【図10】図9の周波数成分に対して単音抽出処理および音源種類仮説生成処理を行って得られた単音ノードを示す図。

【図11】図10の単音ノードに対して、単音グラフ作成処理および音源種類仮説選択処理の中のステップ503までの処理を行った結果を示す図。

【図12】単音グラフ作成の説明図。

【図13】テスト曲に対する音源同定実験結果を同定精度として示した図。

【図14】この発明方法による処理精度を評価する実験に用いた楽譜を示す図。

【図15】前記実験結果を楽譜形式で示した図で、Aは従来法、Bはこの発明方法による図である。

【図1】

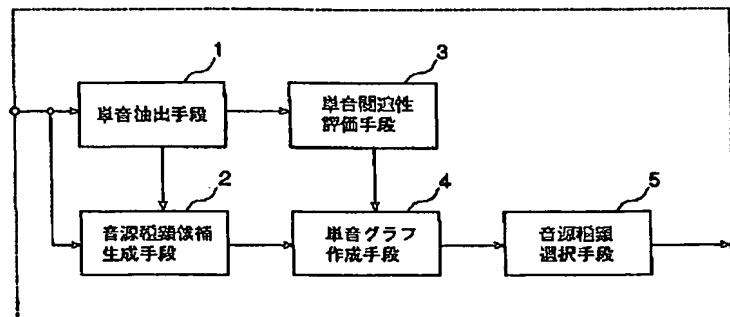


図 1

【図2】

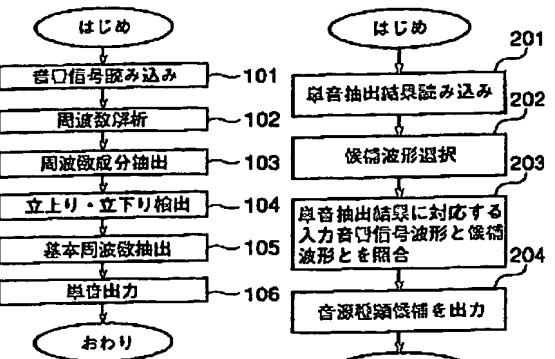


図 2

【図3】

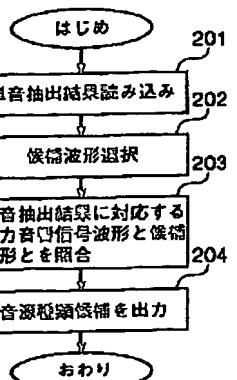


図 3

【図4】

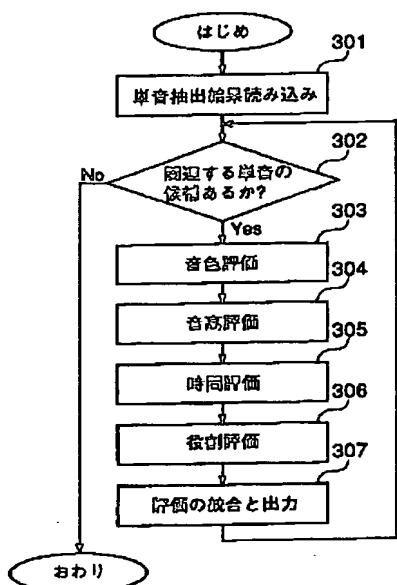


図 4

【図5】

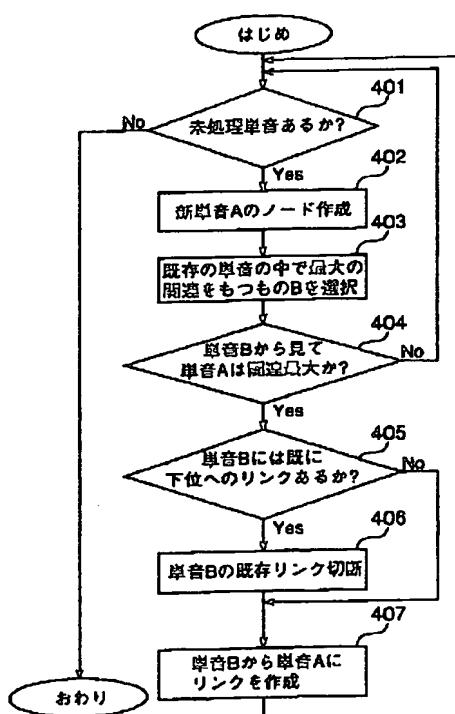


図 5

【図6】

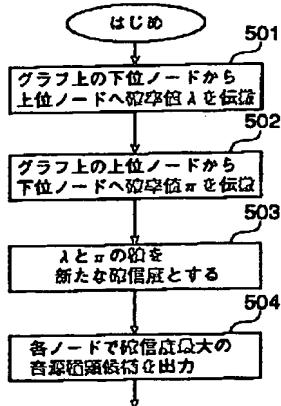


図 6

【図12】

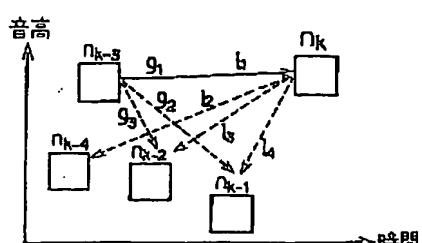
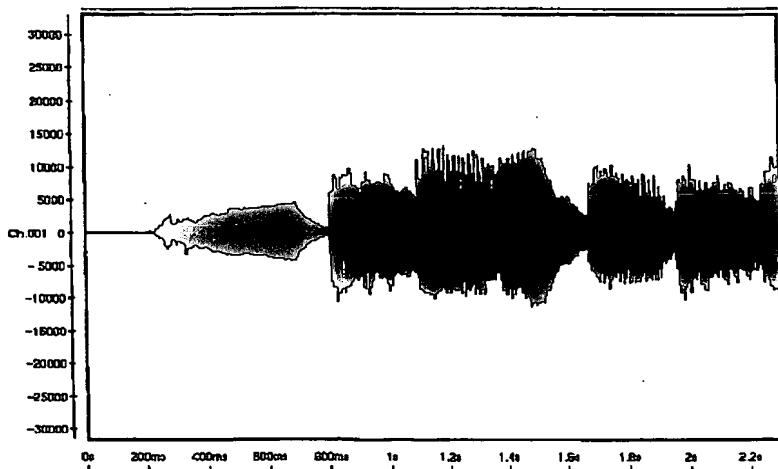
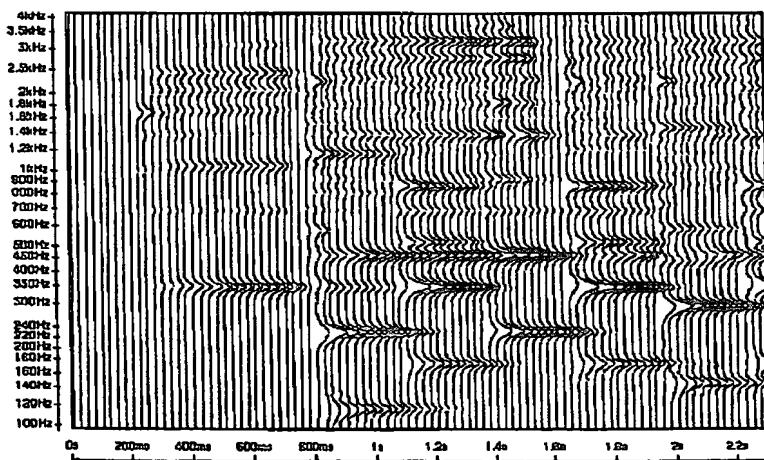


図12

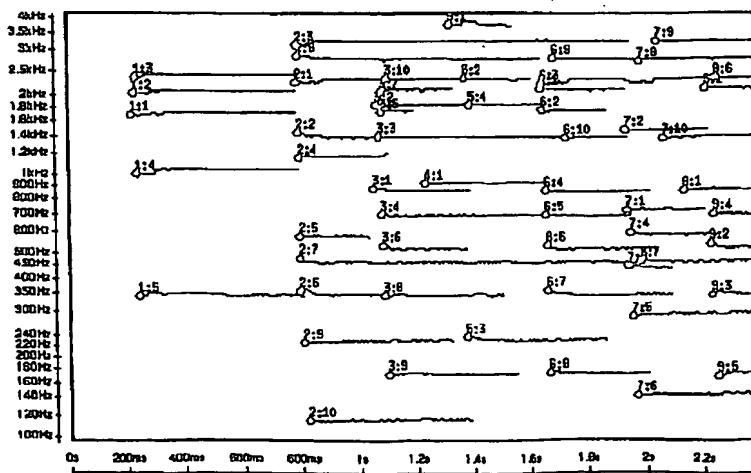
【図7】



【図8】



【図9】



【図13】

固定精度[%]

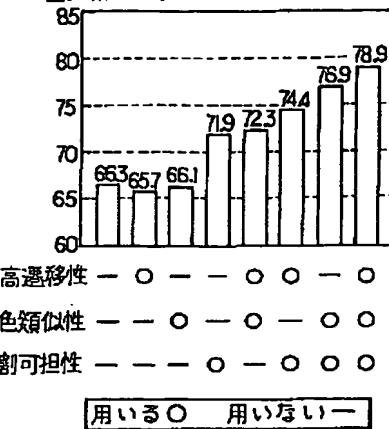


図7

図13

図8

図9

【図10】

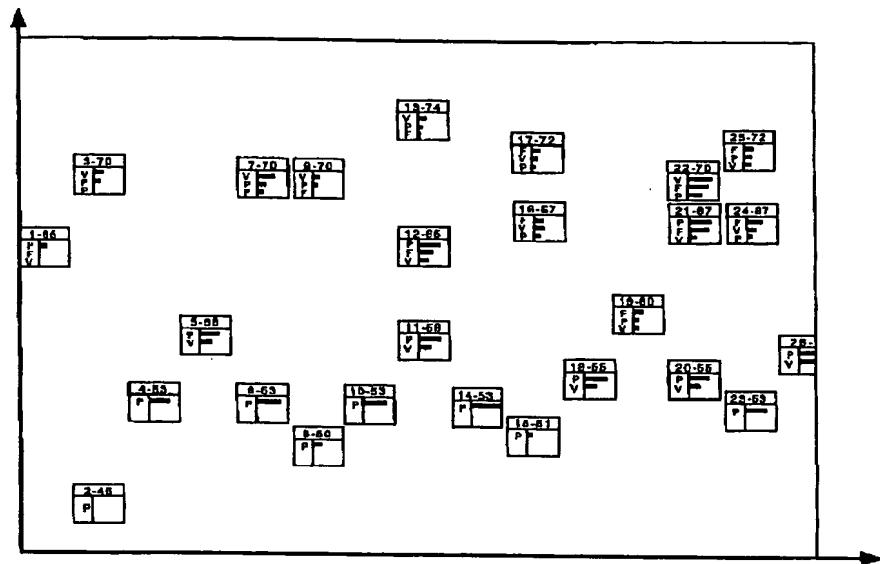


図10

【図11】

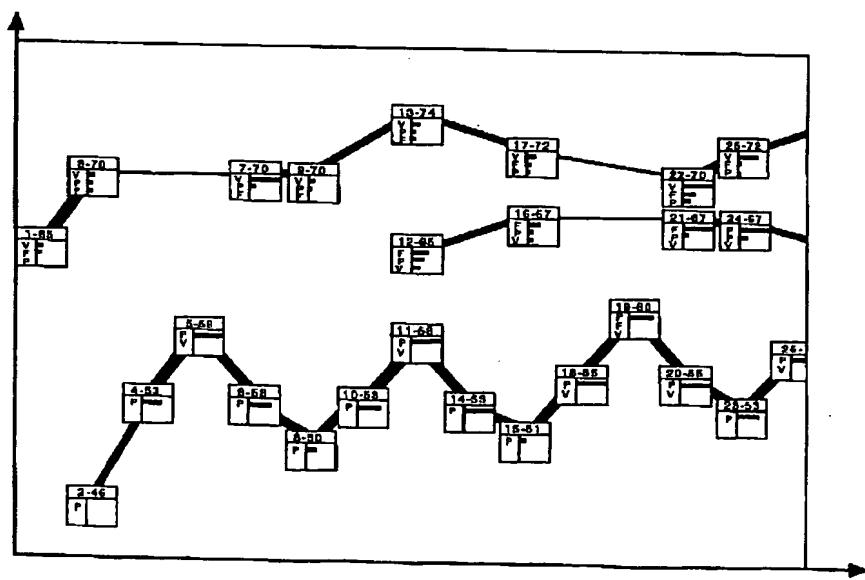


図11

【図14】

Auld Lang Syne
Scotch Traditional

B♭ B♭ Cm F9 Gm B♭ Cm E♭
B♭ B7 Cm Cm B♭ F9 B♭ E♭
B♭ B7 Cm Cm Dm B7 E♭ Cm
B♭ B7 Cm Cm B♭6 F9 B♭ B♭

図14

【図15】

図15A

Part 1
Part 2
Part 3

D E F# G A B C# D#

図15B

Part 1
Part 2
Part 3

D E F# G A B C# D#

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.